Time-dependent changes in mortality and transformation risk in MDS

Pfeilstocker, Michael; Tuechler, Heinz ; Sanz, Guillermo ; Schanz, Julie ; Garcia-Manero, Guillermo; Solé, Françesc

Published in:
Blood

DOI:
10.1182/blood-2016-02-700054

Publication date:
2016

Document Version
Peer reviewed version

~~**DIFFERING TIME DEPENDENT CHANGES IN MORTALITY AND TRANSFORMATION RISK IN MYELODYSPLASTIC SYNDROMES SUBGROUPS**~~

TIME-DEPENDENT CHANGES IN MORTALITY AND TRANSFORMATION RISK IN MDS

Michael Pfeilstöcker[1], Heinz Tuechler[2], Guillermo Sanz[3], Julie Schanz[4], Guillermo Garcia-Manero[5], Francesc Solé[6], John M. Bennett[7], David Bowen[8], Pierre Fenaux[9], Francois Dreyfus[10], Hagop Kantarjian[5], Andrea Kuendgen[11], Luca Malcovati[12], Mario Cazzola[12], Jaroslav Cermak[13], Christa Fonatsch[14], Michelle M. Le Beau[15], Marilyn L. Slovak[16], Alessandro Levis[17], Michael Luebbert[18,] Jaroslaw Maciejewski[19], Sigrid Machherndl-Spandl[20], Silvia M. M. Magalhaes[21], Yasushi Miyazaki[22], Mikkael A. Sekeres[19], Wolfgang R. Sperr[23], Reinhard Stauder[24], Sudhir Tauro[25], Peter Valent[26], Teresa Vallespi[27], Arjan A. van de Loosdrecht[28], Ulrich Germing[11], Detlef Haase[4] and Peter L. Greenberg[29]

(1) Hanusch Hospital and L.Boltzmann Cluster Oncology, Vienna, Austria;

(2) L. Boltzmann Institute for Leukemia Research, Vienna, Austria;

(3) Hospital Universitario La Fe, Valencia, Spain

(4) Georg August Universität, Göttingen, Germany;

(5) The University of Texas, MD Anderson Cancer Center, Houston, TX

(6) Hospital del Mar, Barcelona, Spain;

(7) James P. Wilmot Cancer Center, University of Rochester Medical Center, Rochester, NY;

(8) St James's University Hospital, Leeds, United Kingdom;

(9) Hopital Avicenne, Assistance Publique–Hopitaux de Paris/University Paris XIII, Bobigny, France;

(10) Hopital Cochin, AP-HP University of Paris V, Paris, France;

(11) Heinrich-Heine University Hospital, Düsseldorf, Germany;

(12) Fondazione Istituti di Ricovero e Cura a Carattere Scientifico Policlinico San Matteo and University of Pavia, Pavia, Italy

(13) Institute of Hematology and Blood Transfusion, Praha, Czech Republic

(14) Medical University of Vienna, Vienna, Austria

(15) University of Chicago Comprehensive Cancer Research Center, Chicago, IL;

(16) Quest Diagnostics Nichols Institute, Chantilly, VA

(17) Antonio e Biagio e C Arrigo Hospital, Alessandria, Italy

(18) University of Freiburg Medical Center, Freiburg, Germany

(19) Cleveland Clinic, Cleveland, OH

(20) Elisabethinen Hospital, Linz, Austria

(21) Federal University of Ceara, Fortaleza, Brazil

(22) Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

(23) Division of hematology, Medical University of Vienna, Vienna, Austria

(24) University Hospital of Innsbruck, Innsbruck, Austria

(25) University of Dundee, Scotland, United Kingdom

(26) Division of hematology, Medical University of Vienna and L.Boltzmann Cluster Oncology, Vienna, Austria;

(27) Hospital Universitario Vall d'Hebron, Barcelona, Spain

(28) VU University Medical Center, Amsterdam, The Netherlands

(29) Stanford Cancer Institute, Stanford, CA

Acknowledgments

This work was in part supported by the MDS Foundation, Inc.

Correspondence to:
Prof. Michael Pfeilstöcker, MD, MBA
3rd. Med. Dept. Hanusch Hospital
H. Collinstr.30
1140 Vienna, Austria
Tel.: + 43 1 91021 - 85522
Fax.: + 43 1 91021 -85529
E-Mail: michael.pfeilstoecker@wgkk.at

Running head

TIME-RELATED CHANGE OF RISK IN MDS SUBGROUPS

word counts:

text: 3342

abstract: 245

figure/table count: 1 table, 5 figures,

1 supplement contains 1 table, 2 fig, 1 text

reference count: 25

scientific category: Clinical Trials and Observations

Key points:

Hazards regarding mortality and AML transformation in MDS diminish over time in higher risk, remain stable in lower risk patients.

This change of hazard indicates time-dependent attenuation of power of basal risk scores, relevant for clinical decision making.

Abstract

In myelodysplastic syndromes (MDS) evolution of risk for disease progression or death has not been systematically investigated despite being crucial for correct interpretation of prognostic risk scores. In a multicenter retrospective study we describe changes in risk over time, the consequences for basal prognostic scores and their potential clinical implications. Major MDS prognostic risk scoring systems (IPSS, IPSS-R, WPSS, LR-PSS) and their constituent individual predictors were analyzed in 7,212 primary untreated MDS patients from the IWG-PM database. Changes in risk of mortality and of leukemic transformation over time from diagnosis were described. In higher risk MDS, hazards regarding mortality and AML transformation diminished over time from diagnosis, whereas they remained stable in lower risk patients. After approximately 3.5 years, hazards in the separate risk groups became similar and essentially equivalent after five years. This fact led to loss of prognostic power of different scoring systems considered – more pronounced for survival. Inclusion of age resulted in increased initial prognostic power for survival and less attenuation in hazards. If needed for practicability in clinical management the differing development of risks suggested a reasonable division into lower and higher risk MDS based on the IPSS-R at a cut-off of 3.5 points. Our data regarding time-dependent performance of prognostic scores reflect the disparate change of risks in MDS subpopulations. Lower risk patients at diagnosis remain “constant lower risk", while "initially high risk" patients demonstrate decreasing risk over time. This change of risk should be considered in clinical decision making.

Introduction

The myelodysplastic syndromes (MDS) are clonal hematopoietic stem cell disorders resulting in ineffective hematopoiesis in the bone marrow associated with peripheral blood cytopenias and a risk of developing acute myeloid leukemia (AML) in approximately one third of cases[1]. Due to its heterogeneous nature and wide range of clinical courses, classification and prognostication is of paramount importance for the management of this disease. Independent prognostic instruments have been developed over the years[2] and function as additional staging systems. Most scoring systems assess prognosis at the time of diagnosis assuming stable predictability over the disease course. An earlier single center study has shown moderate loss of prognostic power over time in scoring systems using clinical parameters, whereas systems focusing on cytogenetics and comorbidity maintained prognostic power[3]. Recently, new prognostic scoring systems have been developed providing improved prognostication for MDS patients[4-7]. Previously, the comparison of scores was only provided from the time of diagnosis[7] and the stability of risk over time and the clinical applicability for time points after diagnosis remain unclear.

The aim of this multicenter retrospective study therefore was to assess the relative stability of the newly developed scoring systems over time, to compare the observed time-related changes in prognostic power among these systems and to relate these changes to the time-dependence of hazards. These data can then be applied to MDS populations with different risk for their clinical outcomes when evaluated over time after diagnosis and as they relate to their clinical implications.

Patients and Methods

Patients

This study is based on 7,212 untreated (i.e. they did not receive disease modifying treatment during MDS-phase - disease-specific treatment was allowed after progression to AML) primary MDS patients from 19 institutional databases comprising the International Working Group for Prognosis in MDS (IWG-PM), which generated the Revised International Prognostic Scoring System (IPSS-R) for MDS[7] under the aegis of the MDS Foundation, Inc. in accordance with institutional review board approvals. Patient characteristics were well comparable with other MDS series: median age 71 years, male gender 60%. After a median follow up time of 4.0 years, median overall survival was 3.8 years (range 0.1-39.75, 95% confidence interval 3.7-4.0) and median time to transformation to AML was not reached with 25% of patients transforming to AML after 6.8 years. Patients were diagnosed and classified by French American and British (FAB) and/or World Health Organization (WHO) morphologic classifications; the cytogenetic pattern was classified by original IPSS subtypes[8] and by the refined proposal that was integrated into the IPSS-R[9].

Investigated parameters and scoring systems

First, single score components [hemoglobin, absolute neutrophil counts (ANC) and platelet count, bone marrow blast percentage, cytogenetics] and differentiating features for the IPSS-R (age, performance status, ferritin, LDH, ß-2microglobulin, marrow

fibrosis)[7] were analyzed for stability of prognostic power over time. Then the following scoring systems were analyzed: IPSS[8], IPSS-R and IPSS-RA (i.e. IPSS-R including age)[7], the original WPSS applying transfusion need[4], its modification using hemoglobin thresholds ("WPSS 2011")[10] as well as its modification including age ("WPSS-A")[11], and the low-risk MD Anderson Score (LR-PSS)[5]. For a detailed description and distribution of variables analyzed, see Table 1 and Supplement from reference 7. In the present analysis the IPSS-R[7] was calculated for all cases (n=7212), whereas the WPSS-R[11] was applicable only to patients classified by WHO criteria (n=5763) (i.e. missing a portion of cases present in the IPSS-R). Therefore, when comparing scores including WPSS, only WHO-classified patients were analyzed.

Statistical Methods

Time variations were described by the Cox-zph-test[12], and by applying Dxy[13], a measure of concordance for censored data, at different observation periods (landmarks). Smoothed hazard graphs for time intervals were calculated[14]. Cause-specific hazard was estimated for time to leukemic transformation, and for consistency the related Kaplan-Meier curves are based on these cause-specific hazards, i.e. they do not account for competing risks of death, as the aim was to describe the change of hazard with time, not the description of the cumulative incidence of transformations[15]. The potential influence of scoring weights on stability was analyzed by the creation of hypothetical time-specific scores as landmark reference scores.

The following specific hypotheses connected to the interpretation of results required particular methods: Most scores are based on Cox-models. In these models, a certain set of potential predictors is selected, and the model estimates optimal weights, to form

a prognostic score as a weighted sum of these potential predictors. Since the majority of events occurred early, the estimation was dominated by the initial period after diagnosis. Our question of interest was to determine whether weights, optimized for later periods would improve prediction in these later periods. Therefore, hypothetical scores were estimated for the landmarks 12 to 48 months.

Since the aim of this project did not imply clinically plausible null hypotheses, formal significance tests were not considered appropriate. All analyses were performed using the open source software R version 3.1.2 [16], including the packages "survival"[17], and "bshazard"[14]. For an additional discussion of methods and models used, see the Supplemental methodological material.

Results

Changes in the subgroup specific hazards over time

Changes of hazard over time are shown by smoothed hazard plots, Figure 1A, using IPSS-R for WHO classified patients, serving as an example. Smoothed hazard plots basically contain the same information as the corresponding Kaplan-Meier curves (Figure 1B) The information is displayed in a different form to better visualize changes in risk at different time intervals.

The enlarged section of the survival curves by IPSS-R (Figure 1B) shows, for example, that the curve for "very high" risk starts at 1.0 (i.e., 100%) (at two months because stable disease for two months was one of the inclusion criteria for the IPSS-R) and declines to about 0.9 (i.e., ca. 90%) after one month **(i.e., after 3 months from diagnosis).** That

means the initial mortality is 10% per month. The smoothed hazard for "very high" risk roughly indicates 0.1 (i.e. 10%) hazard (roughly interpretable as 10% mortality per month) in the beginning (see top arrow in Figure 1A). This is consistent with data represented by the Kaplan-Meier-curve.

However, after 30 months the hazard plot (Figure 1A, middle arrow) shows .05, i.e. a 5% monthly mortality for the very high-risk group. While a Kaplan-Meier-curve shows the estimated proportion of persons still alive at each point in follow up time, the hazard plot shows the estimated proportion dying in a defined interval (here one month), given that a person is still alive at the start of the interval of interest. The fact that the force of mortality in the very high risk group decreases from 10% per month to 5% per month after about 2.5 years is clearly visible in the hazard plot, but not readily seen in the survival curve.

For the entire sample (black dashed line, lower arrow), the hazard plot shows 2% mortality per month after diagnosis, and about 1% after 120 months. It can be seen that the mortality risks of the remaining patients for all risk groups are similar after about 60 months. The graph illustrates that similarity of risks derives mainly from a decline in the higher risk groups (IPSS-R very high and high), whereas the mortality risk in the lower risk groups (IPSS-R low and very low) remains essentially unchanged.

In Figure 2, the typical Kaplan-Meier plots are shown adjacent to hazard plots to facilitate comparison and to visualize the proportion of concerned patients at different time points. For a detailed view of individual scores, enlarged single figures are available in the supplemental material (Supplemental figure 1, A-T). The prognostic power of a score, as measured in this project by Dxy, results from the differences in the hazards of

risk categories. Consequently, the attenuation of these hazards with time reduces the time-specific prognostic power of the score. The more similar the hazards of risk categories are from a specific time point onwards, the less prognostically informative is the original assignment to a category.

Change in the subgroup specific hazards over time for prognostic scores

Attenuation of hazards over time was evident for all scoring systems. After approximately 3.5 years, hazards in the separate risk groups become similar and essentially equivalent after five years (see hazard plots in Figure 2). Almost all scores similarly lost prognostic power over time from diagnosis, so that the relative ranking remained virtually unchanged (Figure 3). Scores with high initial prognostic power, even if decreasing over time, retained a greater prognostic capacity than initially weaker scores (see Table 1). Compared to evaluation of survival, we observed a weaker decline of prognostic power with respect to time to transformation to AML.

IPSS-R vs IPSS and FAB vs WHO

Figure 3 demonstrates that the prognostic power of the IPSS-R after about 9 months remains as high as that of the original IPSS at diagnosis for both endpoints. Similarly, the IPSS-RA maintains a comparable power until around 14 months after diagnosis (see gray arrows in Figure 3). The IPSS-R performs generally better if the patient sample is restricted to WHO-defined-MDS (excluding oligoblastic AML) (see Table 1, line “IPSS-R WHO only”).

WPSS

The WPSS variants show high initial prognostic power with loss in power over time, similar to that of other scoring systems. A meaningful comparison with the IPSS-R, necessarily based on WHO classifiable patients only, demonstrates similar initial high prognostic power and attenuation over time.

Inclusion of age in prognostic scoring systems

Inclusion of patient´s age results in higher initial prognostic power and better stability in predicting survival but not for time to AML progression (see comparison of scores with and without age adjustment in Table 1 and Figure 3A).

Stability of scores in lower vs higher risk patients

Scores applied to lower risk MDS only show generally lower prognostic power (because of less risk variation), but remain more stable over time (see Table 1, lines “LR-PSS in LR” and “ IPSS in LR”), because they are less affected by the attenuation of subgroup specific hazards (Figures 4B, 4D). This is seen for both the LR-PSS - initially derived from IPSS low and intermediate-1 patients[5] - and for the IPSS-RA, if restricted to the very low/low/intermediate risk patients according to IPSS-R (see Figures 4A, 4C for survival curves). This ad hoc definition was used only for affording a better comparison of the predictive power of the IPSS-R and the LR-PSS. In contrast, “high/very high risk categories” show a sharp decline in risk over time (e.g., IPSS-R in Figure 2 F, H) for both endpoints. In lower risk MDS, a slight increase of mortality risk occurs due to age (Figures 4C, 4D), whereas no similar effect is observed for time to transformation (Figures 2 K, 2L).

Decline of prognostic power over time in potential prognostic variables

Decline of prognostic power for single potential prognostic variables as measured by Dxy is tabulated in Table 1. Bone marrow blast percentage is the strongest single predictor. Its predictive loss over time is in line with most other parameters and scores. Cytogenetic pattern is of high importance at time of diagnosis with steady loss of prognostic power. Beta-2 microglobulin and performance status (only available for a subset of patients) seem to have moderate, but stable influence on survival and increasing influence on time to progression. As a plausibility check of the analysis**,** conversely, age showed growing negative impact on survival, although not on time to AML progression.

Improvement by hypothetical time-specific scores

Given that most potential prognostic variables exhibit loss of prognostic power over time, we investigated the effect of assigning different weights to score-constituting components for time-specific Cox models. See Supplemental Figure 2 for weights (A, B) and resulting curves (C, D). Dxy values for hypothetical scores for the landmarks 1 to 4 years (overall survival) are shown in Supplemental Table 1. These landmark reference scores do not show higher prognostic power for target times, but rather attenuation of prognostic power.

Consequences for dichotomization into lower risk versus higher risk MDS

Based on prognostic power and on the differing declines of hazards in IPSS-R categories, the optimal dichotomization into just two risk categories – as usually employed in clinical practice – lower risk versus higher risk MDS patients – is the division obtained by using an IPSS-R score of ≤3.5 vs >3.5 points as a cut off.

At time of diagnosis, the Dxy for dichotomization at this cut off is 0.33 for overall survival and 0.44 for time to transformation, respectively, versus 0.31 and 0.37 if cutting at IPSS-R very low/low/intermediate versus high/very high risk groups. Loss of prognostic power at different time points for both approaches can be seen in Table 1. Figure 5, which indicates the respective Kaplan-Meier curves for overall survival (A) and respective hazards plots (B) for patients with an IPSS-R score of ≤3.5 vs >3.5 points, regarding prognostic power also clearly demonstrates this dichotomy, showing good separation of patients with initially higher but declining risk versus patients with constant lower risk. The lower risk group proposed now includes a smaller proportion of cases than the original IPSS lower risk group, consequently assigning more patients to the higher risk group.

Discussion

MDS have been described as a spectrum of dynamic disorders, where clonal evolution identified at a cytogenetic and molecular level may trigger progression[18-19]. On the other hand, time of disease progression may be heterogeneous[1]. Risk for progression and survival may be estimated by risk-based categorization, which is generally performed at the time of diagnosis[2]. Recently, prognostic scoring systems have been substantially improved by refined inclusion and addition of parameters based on the use of larger, comprehensive databases. In addition, their use also for treatment decisions is recommended in disease management guidelines[20-21]. However, data regarding the stability of risk scores over time are scarce[3].

When using prognostic risk scoring systems in clinical daily practice, we assume implicitly that the risk within a specific risk category remains constant over the entire course of the disease. However, the present study showed that all the scoring systems we evaluated for MDS risk categories discriminated better at time of diagnosis compared to later time points (Figure 3, Table 1). In detail, our data demonstrated significant time-dependent changes in the risk for both overall survival and leukemic transformation during elapsing time, which differed for specified patient populations (Figure 2). For higher-risk patients, the mortality risk declined more sharply over time, approaching that observed in lower-risk MDS, whereas for lower-risk patients the mortality risk remained essentially unchanged during follow-up.

One potential explanation for this finding is, that the risk for AML transformation, which has a greater impact in the high risk subset of patients, decreases more dramatically compared to overall survival (Figure 2 F, H). In contrast, age-related risk of death, which has a more pronounced effect on mortality in lower-risk patients, increased with time (Table 1). Risk attenuation is at least partially caused by selective loss of higher-risk patients over time. This loss may be an event such as death or leukemic transformation. In addition, a selection bias cannot be ruled out: the IWG-PM data set used for developing the IPSS-R only includes untreated patients and it is most likely that a greater proportion of higher-risk compared to lower-risk patients is actively treated with passing time from diagnosis, and thus selectively excluded from this study. As changes in risk over time may stem from disease-specific and also from patient-related factors such as comorbidities, loss of prognostic power over time may be not due to a poor

quality of the scoring systems analyzed but to inherent survival dynamics of the MDS patient population.

Theoretically, loss of prognostic power could also result from suboptimal weights of score components. We ruled out this possibility by providing hypothetical scores, optimized for later time intervals comparable to landmark analyses, and observed that these hypothetical scores lost prognostic power in a comparable fashion (Supplemental Table 1). In addition, as single parameters decline to a similar extent, better weighting of these variables at time of diagnosis did not mitigate the decline of prognostic power of scores over time (Table 1). Data using somatic mutational molecular parameters with potential for prognosis[18,21,22] will be of much interest for future analysis. The potential stability of such mutations over time requires further study.

As loss of stability in prognostic scores exists, the question of the best approach for patient re-evaluation arises. Conventionally, the time of first bone marrow examination with features defining MDS is accepted as the time of diagnosis[23]. On the assumption that the impact of features changes little, and remembering different stability of hazards in different risk categories, scoring systems such as the IPSS-R may be used for re-evaluation after diagnosis, at time points where still clinically meaningful conclusions can be drawn from risk score categories. Re-evaluation could be done at specific intervals established according to clinical needs (e.g., every six months for higher risk patients and at longer intervals for lower risk patients).

Our data on the differential development of risks for the single score subgroups suggest also for the first time a reasonable division into lower risk and higher risk MDS into two groups based on the IPSS-R, a fact that may be valuable for design of clinical trials and

for patient management. As described, lower risk MDS have virtually constant risk for both endpoints, whereas in higher-risk cases risk diminishes substantially over time. If the aim is optimal prognostic separation of lower risk versus higher risk patients, then a dichotomization based on 3.5 scoring points of the IPSS-R raw score (i.e. ≤3.5 vs >3.5) yielded the best results. This also best represents the different changes in risk categories over time (Figure 5, Table 1). In particular, all patients scoring >3.5 points fit well into the higher risk group. The proportion of lower risk patients using this dichotomization is smaller compared to the lower risk group of the original IPSS, which was heterogenous and contained patients with higher risk.

Our results serve to underscore the ongoing processes occurring during the course of the disease and should help guide clinical decisions in MDS. One interpretation relates to a new perception of higher and lower risk MDS where lower risk patient subgroups may be better described as a group of patients with “constant lower risk" (risk remaining the same over time) whereas the term "initially higher risk" MDS (but with decreasing risk over time) fits higher risk subgroups better. Although our data are derived from “untreated” MDS patients, this is a representative patient cohort since the majority of MDS patients are still provided with supportive care without disease-modifying treatment[24, 25]. Future analysis of patients receiving treatment and comparison with the data presented herein are planned and may indicate the impact of therapy.

The following clinical recommendations can be derived from our data: If a patient is initially categorized as high risk, options for disease-modifying treatment should be considered immediately, and if deemed appropriate, a decision for treatment should be made as early as possible. If in contrast, due to comorbidities, decreased performance

status, clinical stability or other reasons a high risk patient does not receive disease-modifying treatment at time of diagnosis, but remains in stable condition for a prolonged period, re-evaluation for a specific treatment should be re-considered at that time. For lower risk patients it should be noted that risk remains relatively constant over time, therefore surveillance and ongoing re-evaluation should be maintained long term.

In conclusion, our data describe the change of risk within prognostic score categories over time in MDS and their effects on the construction and interpretation of prognostic scoring systems. Clinicians should be aware of these facts when assessing patients after time intervals and when making treatment decisions. This study clearly demonstrates that a cut-off point of 3.5 in the IPSS-R scoring system is the best for segregating MDS patients into two risk groups – lower risk and higher risk – for therapeutic purposes, although a loss of prognostic power compared to the use of raw score data or the five IPSS-R category approach should be noted. Since the statistical tools used in our analysis may be applied to other prognostic scoring systems, it will be of interest to determine whether similar changes of risk over time are observed in other disease entities.

Contributions:

M.P. designed, performed, and coordinated the research, collected, contributed, analyzed and interpreted the data, and wrote the manuscript; H.T. designed and performed the research, performed the statistical analyses, produced the figures, and edited the manuscript; P.L.G, G.S., A.A.v.d.L collected, contributed, analyzed, and

interpreted data, and edited the manuscript. J.M.B., G.G.-M., F.S., D.B., P.F., A.L., J.C., M.L., J.M., S.M.M.M., S.M., Y.M., M.P., M.S., W.R.S., J.S., R.S., S.T., P.V., T.V., F.D., H.K., A.K., L.M., M.C., D.H and U.G. collected and contributed data, analyzed the results and critically revised the paper; C.F., M.M.L.B., and M.L.S. analyzed and interpreted the data and critically revised the paper;

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Acknowledgments

This work was in part supported by the MDS Foundation, Inc.

References

1. Ades L, Itzykson R, Fenaux P: Myelodysplastic syndromes. *Lancet* 2014; 383(9936),2239–52.

2. Pfeilstöcker, M: Prognostic Scoring in MDS, in Varkonyi J (ed): The Myelodysplastic syndromes. Dordrecht, Heidelberg, London, New York, Springer, 2011, pp 103-120.

3. Pfeilstöcker M, Tüchler H, Schönmetzler A, et al: Time changes in predictive power of established and recently proposed clinical, cytogenetical and comorbidity scores for Myelodysplastic Syndromes. *Leuk Res* 2012; 36(2),132–139.

4. Malcovati L, Germing U, Kuendgen A, et al: Time-dependent prognostic scoring system for predicting survival and leukemic evolution in myelodysplastic syndromes*. J Clin Oncol* 2007; 25(23,3503–3510.

5. Garcia-Manero G, Shan J, Faderl S, et al: A prognostic score for patients with lower risk myelodysplastic syndrome. *Leukemia* 2008; 22(3),538–543.

6. Kantarjian H, O'Brien S, Ravandi F, et al: Proposal for a new risk model in myelodysplastic syndrome that accounts for events not considered in the original International Prognostic Scoring System. *Cancer* 2008;113(6),1351–1361.

7. Greenberg PL, Tuechler H, Schanz J, et al: Revised international prognostic scoring system for myelodysplastic syndromes. *Blood* 2012; 120(12), 2454–2465.

8. Greenberg P, Cox C, LeBeau MM, et al.: International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* 1997; 89(6), 2079–2088.

9. Schanz J, Tuechler H, Sole F, et al: New comprehensive cytogenetic scoring system for primary myelodysplastic syndromes (MDS) and oligoblastic acute myeloid leukemia after MDS derived from an international database merge. *J Clin Oncol* 2012;30(8), 820–829.

10. Malcovati L, Della Porta MG, Strupp C, et al: Impact of the degree of anemia on the outcome of patients with myelodysplastic syndrome and its integration into the WHO classification-based Prognostic Scoring System (WPSS). *Haematologica* 2011; 96(10), 1433–1440.

11. Della Porta MG, Tuechler H, Malcovati L, et al: Validation of WHO classification-based Prognostic Scoring System (WPSS) for myelodysplastic syndromes and comparison with the revised International Prognostic Scoring System (IPSS-R). A study of the International Working Group for Prognosis in Myelodysplasia (IWG-PM). *Leukemia* 2015; 29(7),1502-1513.

12. Grambsch PM, Therneau TM: Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; 81(3),515–526.

13. Harrell Jr, Frank E., and with contributions from many other users: Hmisc: Harrell Miscellaneous (version R package version 3.8-2). 2010; http://CRAN.R – project.org/package=Hmisc.

14. Rebora, P., Salim, A., Reilly, M: bshazard: A Flexible Tool for Nonparametric Smoothing of the Hazard Function. R J 6:114–22, 2014

15. Gooley TA, Leisenring W, Crowley J, Storer B: . Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med 1999;18, 695–706.

16. R Core Team. R: A Language and Environment for Statistical Computing. Vienna Austria R Found Stat Comput: http://www.R-project.org, 2014

17. Therneau, Terry M. A Package for Survival Analysis in S. http://mayoresearch.mayo.edu/mayo/research/biostat/upload/survival.pdf., 1999

18. Papaemmanuil E, Gerstung M, Malcovati L, et al: Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 2013;122(22),3616–3627.

19. Jabbour E, Takahashi K, Wang X, et al: Acquisition of cytogenetic abnormalities in patients with IPSS defined lower-risk myelodysplastic syndrome is associated with poor prognosis and transformation to acute myelogenous leukemia. *Am J Hematol* 2013; 88(10), 831–837.

20. Malcovati L, Hellstrom-Lindberg E, Bowen D, et al: Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. *Blood* 2013; 122(17),2943–2964.

21. Greenberg PL, Stone RM, Bejar R, et al: Myelodysplastic Syndromes, Version 2.2015. *J Natl Compr Canc Netw* 2015;13(3), 261–272.

22. Malcovati L, Karimi M, Papaemmanuil E, et al: SF3B1 mutation identifies a distinct subset of myelodysplastic syndrome with ring sideroblasts. *Blood*. 2015; 126(2), 233-41

23. Valent P, Horny HP, Bennett JM, et al: Definitions and standards in the diagnosis and treatment of the myelodysplastic syndromes: Consensus statements and report from a working conference. *Leuk Res* 2007; 31(6),727–736.

24. Pease DF, Ross JA, Poynter JN, et al: Differences in community and academic practice patterns for newly diagnosed myelodysplastic syndromes (MDS) patients. *Cancer Epidemiol.* 2015; 39(2), 222–228.

25. Kelaidi C, Stamatoullas A, Beyne-Rauzy O, et al: Daily practice management of myelodysplastic syndromes in France: data from 907 patients in a one-week cross-sectional study by the Groupe Francophone des Myélodysplasies. *Haematologica* 2010; 95(6),892–899.

Legends for figures and table

Figure 1. Survival of IPSS-R classified patient subgroups using smoothed hazard plots (A) and corresponding Kaplan-Meier curves (B) (representative example).

Smoothed hazard plots more clearly demonstrate changes in risk at different time intervals than do Kaplan-Meier plots. The smoothed hazard for "very high" risk indicates 10% monthly mortality risk in the beginning (top arrow in Figure 1A) in agreement with the Kaplan-Meier curve. After approximately 30 months (middle arrow) 5% monthly mortality for the very high-risk group is shown, not clearly visible in the Kaplan-Meier curve. The mortality risks of the remaining patients for all risk groups are similar after approximately 60 months. Note that the time scale in (B) is expanded to improve visibility of the decline in the first year.

vhr = very high-risk, all patients = bold black dotted line.

Figure 2: Comparison of Kaplan-Meier curves and hazard plots for specific risk scoring systems, both for overall survival (columns 1-2) and for time to leukemic transformation (columns 3-4). Colors for risk groups are assigned in the order of risk from lowest to highest: green, gray, yellow, red, blue. All patients: bold black dotted line. For leukemic transformation the cause-specific hazard is shown. The curves for time to leukemic transformation correspondingly are based on the cause-specific hazard (and are not cumulative incidence curves).

Attenuation of hazards occurred over time after diagnosis in all scoring systems. After approximately 3.5 years, hazards in the separate risk groups became similar, and essentially equivalent after five years. **Note differing time scales for the Kaplan-Meier and hazard plots.**

For a detailed view of individual scores, enlarged color single figures are available in the supplemental material (Supplemental Figure 1 A-T)

Figure 3: Comparison of change in prognostic power for specific scoring systems: A for overall survival, B for time to leukemic transformation (based on WHO-classified patients).

The figure demonstrates that nearly all scores lost prognostic power over time, with the relative ranking remaining virtually unchanged. Scores with high initial prognostic power remained prognostically stronger than initially weaker scores. The prognostic power of the IPSS-R after about 9 months and the IPSS-RA until around 14 months remain as high as that of the original IPSS at diagnosis (gray arrows). Inclusion of patient´s age results in higher initial prognostic power and better stability in predicting survival but not for time to AML progression, represented by results from IPSS-R(A) and WPSS(A) versions which included age (in Fig 3A).

Figure 4. Stability of scores in lower risk patients: (Kaplan-Meier curves and hazard plots)

Overall survival shown by Kaplan-Meier curves (A, C), and hazard plots (B, D) using LR-PSS (A, B) and IPSS-RA (C, D). The figure demonstrates that scores applied to lower risk MDS only have lower prognostic power, but remain more stable over time, and are less affected by the attenuation of subgroup specific hazards (Figure 4B, 4D). This is seen for both the LR-PSS and for the IPSS-RA, with both scores restricted to the IPSS-R very low/low/intermediate patients (see Figure 4A, 4C for survival curves). An increase of mortality risk related to age is shown (Figure 4C, 4D). LR-PSS categories: C1: score 0-2, C2: score 3-4, C3: score 5-7. **Note differing time scales for the Kaplan-Meier and hazard plots.**

Figure 5: Dichotomized separation of lower vs higher risk MDS patients in IPSS-R stratified patients: (Kaplan-Meier curves and hazard plots: Kaplan-Meier curves for overall survival (A) and respective hazards plots (B) for patients with an IPSS-R score of ≤3.5 vs >3.5 points (A-B) yielding the best results regarding prognostic power and showing a good separation of patients with initially higher but declining risk vs patients with constant lower risk. **Note differing time scales for the Kaplan-Meier and hazard plots.**

Table 1: Dxy values for specific scores and single clinical predictors

Dxy is a measure of correlation varying between -1 and 1, with 0 indicating no correlation, and 1 perfect concordance of prognosis and survival, respectively time to transformation (see also the short explanation of Dxy in the supplement).

Dxy values were tabulated conditional on minimum observation time for potential predictors and composite scores. Changes in Dxy values are consistent with hazard plots Figures 1-2 and 4-5 and show similar loss of prognostic power over time. Scores with high initial prognostic power, even if decreasing over time, remained prognostically stronger than initially weaker scores. With respect to single parameters, bone marrow blast percentage is the strongest single predictor. Its predictive loss over time is consistent with most other parameters and scores. Cytogenetic pattern is of high importance at time of diagnosis with steady loss of prognostic power. Age showed growing negative impact on survival, although not on time to AML progression.

For hemoglobin, neutrophils, platelets and bone marrow blasts the cut points were those used for the IPSS-R. Age groups were categorized ≤55, >55 to ≤65, >65 to ≤75, >75 to ≤80 and >80. For serum LDH and beta-2 microglobulin the upper limit of normal was the cut point; for serum ferritin 350ng/ml was the chosen cut point. Cytogenetic categories are those used in the IPSS-R. “In LR” denotes application of LR-PSS and IPSS-R on lower risk patients (i.e. IPSS-R very low, low, intermediate) only. IPSS-R WHO: IPSS-R exclusively applied to patients classifiable according to WHO. For dichotomization in two groups with higher vs lower risk patients dxy for combined IPSSR very low, low, intermediate versus high, very high (IPSS-R vlli/hvh) and for a cut off at 3.5 IPSS-R score points (IPSS-R-LH<=3.5/>3.5) are tabulated.

Figures

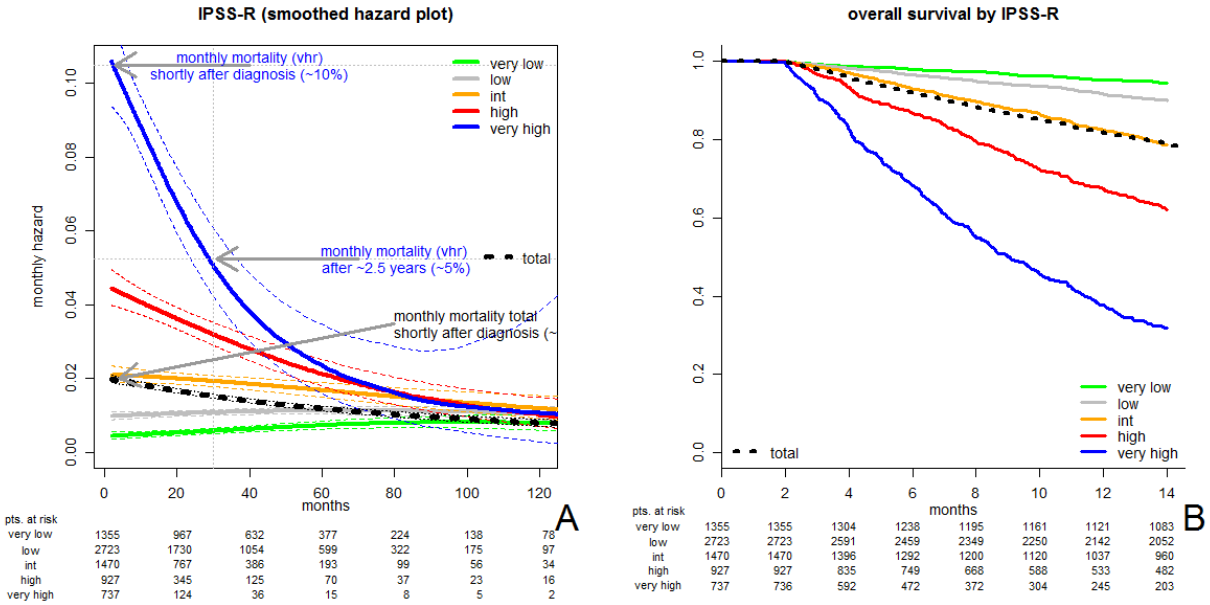

Figure 1

Figure 2

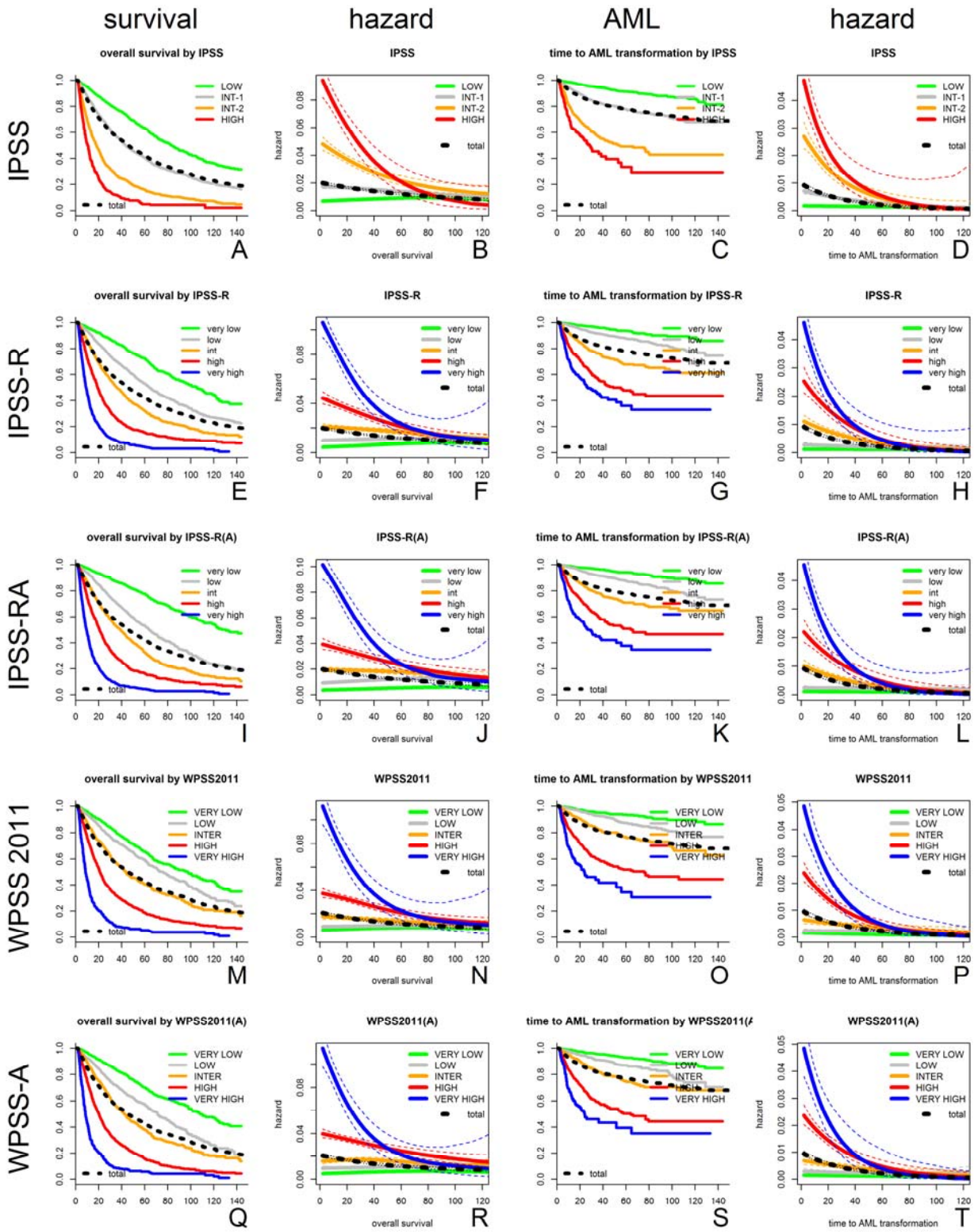
survival
hazard
AML
hazard
IPSS
IPSS-R
IPSS-RA
WPSS 2011
WPSS-A
overall survival by IPSS
IPSS
time to AML transformation by IPSS
overall survival by IPSS-R
IPSS-R
time to AML transformation by IPSS-R
overall survival by IPSS-R(A)
IPSS-R(A)
time to AML transformation by IPSS-R(A)
overall survival by WPSS2011
WPSS2011
time to AML transformation by WPSS2011
overall survival by WPSS2011(A)
WPSS2011(A)
time to AML transformation by WPSS2011(A
LOW
INT-1
INT-2
HIGH
very low
low
int
high
very high
VERY LOW
LOW
INTER
HIGH
VERY HIGH
total
hazard
overall survival
time to AML transformation
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T

Figure 3

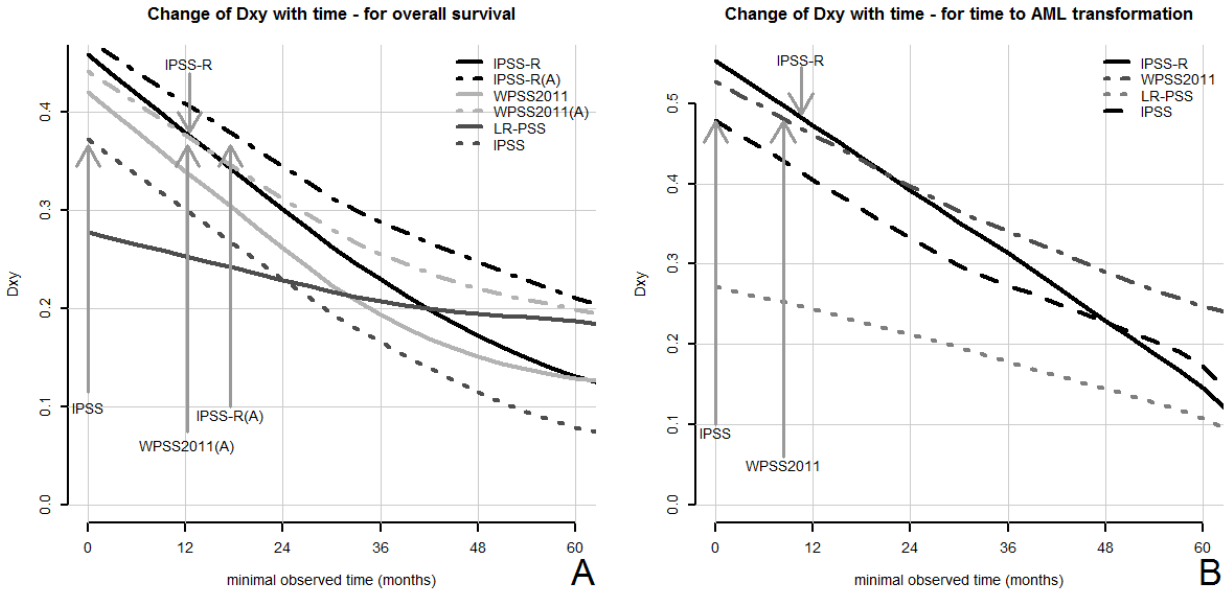
Change of Dxy with time - for overall survival
IPSS-R
IPSS-R
IPSS-R(A)
WPSS2011
WPSS2011(A)
LR-PSS
IPSS
Dxy
0.0
0.1
0.2
0.3
0.4
IPSS
IPSS-R(A)
WPSS2011(A)
0
12
24
36
48
60
minimal observed time (months)
A
Change of Dxy with time - for time to AML transformation
IPSS-R
IPSS-R
WPSS2011
LR-PSS
IPSS
Dxy
0.0
0.1
0.2
0.3
0.4
0.5
IPSS
WPSS2011
0
12
24
36
48
60
minimal observed time (months)
B

Figure 4

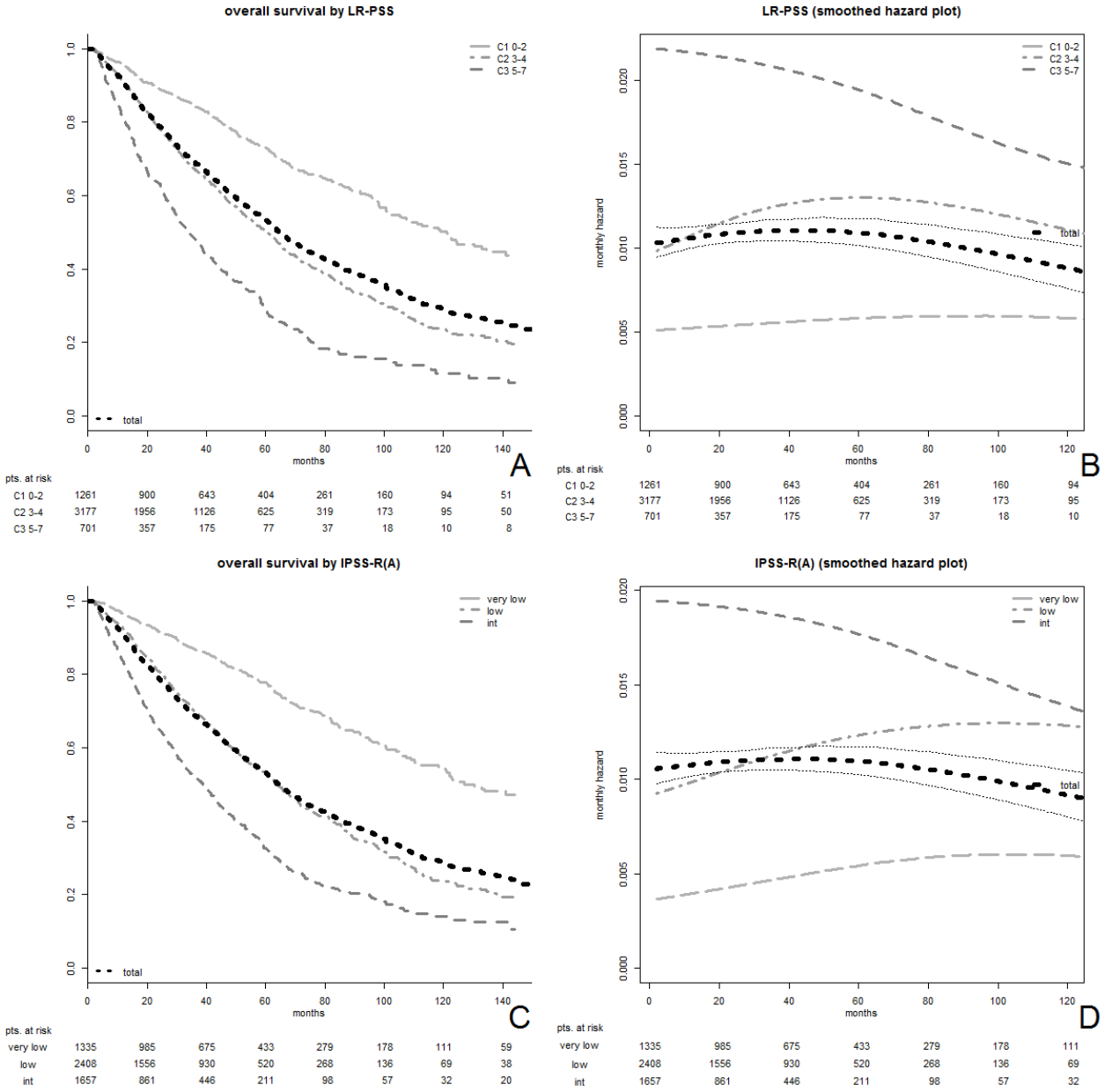
overall survival by LR-PSS
C1 0-2
C2 3-4
C3 5-7
total
months
A
pts. at risk
C1 0-2 1261 900 643 404 261 160 94 51
C2 3-4 3177 1956 1126 625 319 173 95 50
C3 5-7 701 357 175 77 37 18 10 8
LR-PSS (smoothed hazard plot)
monthly hazard
B
pts. at risk
C1 0-2 1261 900 643 404 261 160 94
C2 3-4 3177 1956 1126 625 319 173 95
C3 5-7 701 357 175 77 37 18 10
overall survival by IPSS-R(A)
very low
low
int
C
pts. at risk
very low 1335 985 675 433 279 178 111 59
low 2408 1556 930 520 268 136 69 38
int 1657 861 446 211 98 57 32 20
IPSS-R(A) (smoothed hazard plot)
D
pts. at risk
very low 1335 985 675 433 279 178 111
low 2408 1556 930 520 268 136 69
int 1657 861 446 211 98 57 32

Figure 5

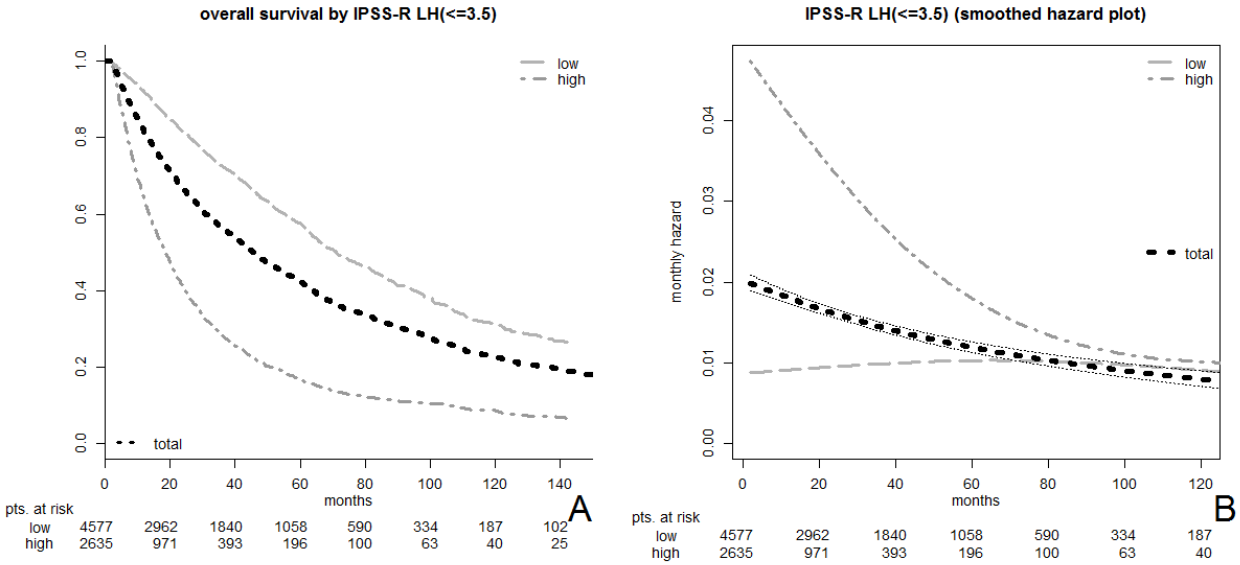
overall survival by IPSS-R LH(<=3.5)
low
high
total
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
120
140
months
pts. at risk
low 4577 2962 1840 1058 590 334 187 102
high 2635 971 393 196 100 63 40 25
A
IPSS-R LH(<=3.5) (smoothed hazard plot)
low
high
total
monthly hazard
0.04
0.03
0.02
0.01
0.00
0
20
40
60
80
100
120
months
pts. at risk
low 4577 2962 1840 1058 590 334 187
high 2635 971 393 196 100 63 40
B

Table 1. Dxy values for specific scores and single predictors

	Dxy values for specified minimum observation times											
	For overall survival						For time to transformation					
Time after diagnosis (months)	0	12	24	36	48	60	0	12	24	36	48	60
Scoring System												
IPSS	0.37	0.30	0.22	0.16	0.11	0.07	0.49	0.38	0.34	0.26	0.25	0.20
IPSS-R	0.43	0.35	0.27	0.20	0.14	0.08	0.53	0.42	0.38	0.31	0.24	0.07
IPSS-RA	0.46	0.38	0.31	0.26	0.22	0.18	0.53	0.43	0.38	0.32	0.25	0.13
WPSS	0.44	0.36	0.29	0.25	0.18	0.21	0.59	0.48	0.39	0.35	0.30	0.22
WPSS 2011	0.41	0.34	0.25	0.18	0.14	0.11	0.53	0.45	0.40	0.34	0.33	0.22
WPSS A	0.44	0.38	0.31	0.24	0.20	0.19	0.52	0.46	0.40	0.33	0.27	0.22
LR-PSS	0.27	0.25	0.25	0.20	0.18	0.19	0.26	0.23	0.24	0.18	0.16	0.17
LR-PSS in LR	0.21	0.20	0.21	0.18	0.18	0.18	0.16	0.18	0.20	0.14	0.12	0.10
IPSS-RA in LR	0.28	0.26	0.25	0.24	0.23	0.19	0.37	0.32	0.28	0.25	0.21	0.09
IPSS-RA WHO only	0.48	0.40	0.34	0.28	0.23	0.20	0.55	0.47	0.40	0.35	0.28	0.15
IPSS-R vlli/hvh	0.31	0.20	0.12	0.06	0.02	0.01	0.37	0.24	0.19	0.13	0.08	0.05
IPSS-R-LH<=3.5/>3.5	0.33	0.26	0.19	0.13	0.07	0.03	0.44	0.33	0.29	0.25	0.19	0.06
Predictor												
Hemoglobin	0.21	0.17	0.16	0.11	0.08	0.05	0.16	0.12	0.12	0.08	0.06	0.08
Neutrophils	0.11	0.09	0.07	0.04	0.03	0.02	0.16	0.15	0.12	0.10	0.05	0.01
Platelets	0.23	0.14	0.09	0.04	0.03	0.04	0.17	0.11	0.07	0.01	-0.01	0.00
Bone marrow blasts	0.30	0.26	0.18	0.13	0.10	0.06	0.48	0.38	0.30	0.27	0.25	0.12
Age	0.09	0.11	0.16	0.19	0.22	0.26	0.04	0.02	0.05	0.03	-0.03	0.16
ECOG	0.16	0.14	0.11	0.06	0.09	0.08	0.10	0.07	0.05	0.12	0.23	0.13
Ferritin	0.15	0.11	0.08	-0.01	0.03	0.01	0.11	0.08	-0.01	0.06	0.01	0.14
LDH	0.12	0.09	0.07	0.04	0.02	0.01	0.12	0.09	0.07	-0.01	0.01	0.08
Beta-2 microglobulin	0.14	0.14	0.14	0.15	0.11	0.10	0.02	0.07	0.04	0.01	0.00	0.38
Bone marrow fibrosis	0.04	0.04	0.05	0.04	0.01	0.02	0.05	0.03	0.09	0.00	0.01	0.05
Cytogenetic categories	0.25	0.14	0.09	0.06	0.02	0.00	0.28	0.19	0.17	0.14	0.06	0.06